Model Based Analysis of Insider Threats

Taolue Chen
Middlesex University London
t.chen@mdx.ac.uk

Tingting Han
Birkbeck, University of London
tingting@dcs.bbk.ac.uk

Florian Kammueller
Middlesex University London
f.krammueller@mdx.ac.uk

Ibrahim Nemli
Technical University Denmark
ibrahimnemli@msn.com

Christian W. Probst
Technical University Denmark
cwpr@dtu.dk

***Abstract*—In order to detect malicious insider attacks it is important to model and analyse infrastructures and policies of organisations and the insiders acting within them. We extend formal approaches that allow modelling such scenarios by quantitative aspects to enable a precise analysis of security designs. Our framework enables evaluating the risks of an insider attack to happen quantitatively. The framework first identifies an insider's intention to perform an inside attack, using Bayesian networks, and in a second phase computes the probability of success for an inside attack by this actor, using probabilistic model checking. We provide prototype tool support using Matlab for Bayesian networks and PRISM for the analysis of Markov decision processes, and validate the framework with case studies.**

I. Introduction

Computer security has always been considered as a cross-cutting problem domain. Today, the real security risks are largely based on the attacker being someone within the organisation; security has to emphasize the top-layer, the organisational part of security. This organisation cannot be reduced to individual interactions between human and computers, but must be viewed from a more global perspective in which "organisation" refers to the infrastructure of a company or other segments of societal life in which humans and physical devices interact.

While attack vectors can be identified through invalidation of policies [4], a *quantitative* analysis is still missing. Moreover, it is crucial that a security analyst in an organisation has tools that enable a concrete estimation of insider risk given an infrastructure, the actors and their profiles. We thus propose in this paper a quantitative framework for the estimation of malicious insider risk that uses probabilities as well as budgets of actors for insider threat analysis.

This is an extended abstract of the paper [3] with the same title, which provides more details on the preliminaries, analysis framework, as well as examples.

II. An Analysis Framework

The overall aim of the analysis is to estimate the probability that an employee of an organisation (conceived as the insider) launches a successful insider attack. Generally speaking, our framework consists of two components: The *intentional analysis* provides a quantitative measure for the risk that a particular employee may reach the tipping point and turn into a malicious insider; the *behavioural analysis* estimates where an insider could successfully launch an attack in a company's infrastructure. This is influenced by the personal characteristics of the attacker, for example, the attacker's skill to break a lock or succeed in social engineering the secretary. We elaborate the *model-based* techniques we propose for the two parts of analysis as follows.

A. Intentional Analysis

Insider threats are usually specified by *System Dynamics* [2]. Our analysis involves enriching the System Dynamics model by quantitative information, *i.e.*, the (conditional) probabilities, giving rise to a Bayesian network. The graph structure of the Bayesian network conforms to the given System Dynamics model. Technically, for each node in the System Dynamics model, we introduce a random variable. In general, we usually have the following two cases:

1) For *events* that might happen, the corresponding random variable is governed by the *Bernoulli distribution*, *i.e.*, it takes value 1 with success probability p and value 0 with failure probability $q = 1 - p$;
2) For quantities of continuous nature, in principle we can introduce a continuous random variable, *e.g.*, the *degree* of dissatisfaction, ranging over $[0, 1]$. However, for computational efficiency, we usually discretise $[0, 1]$ into partitions $[0, 0.1)$, $[0.1, 0.2), \cdots$, or even coarser granularity, such as "low", "mediate", "high", which could correspond to $[0, 0.3)$, $[0.3, 0.7)$, $[0.7, 1]$, respectively.

As the next step, we need to specify the conditional probabilities appearing in the Bayesian network, which can be *learnt* from historical data by standard machine learning techniques.

a) Implementation: We provide a preliminary implementation of the intention analysis framework, by leveraging the *Bayes Net Toolbox* ([7], BNT), a tool for Bayesian network that is an open-source Matlab package for directed graphical models. It supports many kinds of nodes (probability distributions), exact and approximate inference, parameter and structure learning, and static and dynamic models.

We encode our Bayesian network in Matlab in a straightforward way: the directed acyclic graph (given in the System Dynamics model) is specified as an adjacency matrix and the

conditional probabilities are specified by the "CP" constructor provided by BNT.

With the Bayesian network model, one can carry out various analyses, which are typically done by *probabilistic inference*. Below we exemplify two queries by performing exact inference and using only the standard junction tree algorithm (supported by the junction tree engine in BNT).

A typical *likelihood query* for the BN such as

$$\Pr[\text{Steal} = H \mid \text{Den} = 1 \text{ and Event} = 1]$$

can be done by computing joint distributions. Intuitively, this is to compute the probability that the insider's desire to steal is high (Steal $= H$) on the condition that the organisation denies the insider's request (Den $= 1$) and there is a competitor enticing the insider (Event $= 1$). This serves our basic purpose of the intentional analysis.

We can also perform an *a posteriori* query that computes the probability of a precipitating event given that the insider's desire to steal is high, *e.g.*, $\Pr[\text{Event} = 1 \mid \text{Steal} = H]$. This is done by computing marginal distributions. Such an analysis is useful as the analyst can identify the main factors which lead to the insider attack.

B. Behavioural analysis

The second component of our framework is the behavioural analysis of insider threats. Given the infrastructure of the organisation and a personal profile, this analysis estimates the probability of successful insider attacks. This part of the analysis has four steps:

- **Step 1**: Model the infrastructure in the *abstract system specification*.
- **Step 2**: Encode the abstract system specification in ExASyM and generate the *labelled transition system*.
  The transformation performs a reachability analysis to generate the set of valid configurations and transitions. The analysis visits all locations where an actor might be, performs all actions that the actor does not need another data item for, and generates states and transitions accordingly. If a new data item has been obtained or a new location has been reached, the algorithm performs all actions that the new capability enables. This computation is repeated until a fix point is reached.
- **Step 3**: Translate the *transition system* into *Markov decision processes* (MDP) by annotating the transitions with probabilities.
  An MDP for the system can be found in Fig. 1. For the probabilities, we mainly consider: (1) *entering probability*, which, for each state (being physical or logical), specifies the probability for an actor (insider) to access that location by any means. (2) *successful probability*, which, for each terminal state, specifies the probability that an actor manages to accomplish the attack after entering the terminal location. (3) detecting probability, which, for each state, specifies the probability of being caught when the insider attempts to perform the attack.

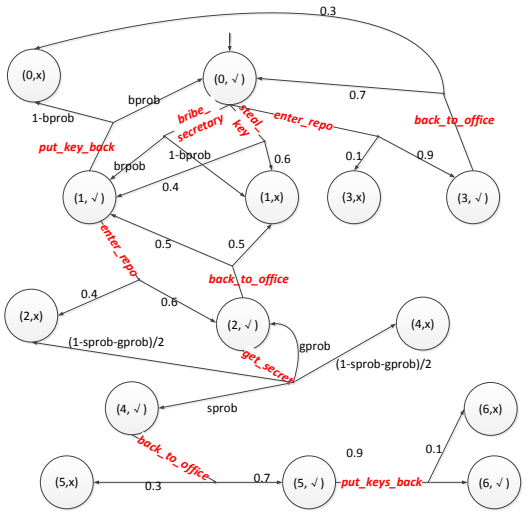

Fig. 1. MDP model of the example

We use the probabilistic model checker PRISM [5] to specify the MDP.

- **Step 4**: Perform behavioural analysis by verification and synthesis of the Markov decision process.
  Now we analyse the obtained MDP by standard probabilistic model checking techniques to compute the maximum probability that the insider steals a confidential file without being caught. Note that the insider has different strategies, for instance, selecting where to go from a physical location, or trying to social engineer other actors. The insider's goal is to maximise success probability, whereas the organisation must consider a worst-case scenario.
  Such a problem boils down to computing the maximum probability to reach the state $(\mathsf{succ}, \checkmark)$ (state $(6, \checkmark)$ in the running example). For this purpose, we consider the query `Pmax=?[F (loc=succ and tick=1)]` where PRISM returns the maximum probability, as well as the corresponding strategy for the insider to achieve this probability.
  By such an analysis, the organisation can identify the potential weakness of the infrastructure, and carry out necessary security improvement.

C. Budget Analysis

The behaviour analysis focuses exclusively on analysing the best strategy of an insider to maximise the chance of success. A tacit assumption is that the insider is determined to launch the insider attack at all costs. However, in practice, typically each (insider) action comes with a *cost*, in terms of money, time, or computation power. The cost of attacks is an important factor when selecting the strategy to launch the attack, or to decide whether such an attack pays off.

Therefore, a refined analysis – *budget analysis* – should take the limited resource of insiders into account. One typical scenario is that the insider has a certain budget to launch the attack. Intuitively, the budget captures the capital the insider

can afford or is willing to pay. Under these limited resources, we intend to identify the maximum probability of success by exploiting MDP models with costs/prices. In other words, each state is equipped with a nonnegative real number (in practice, usually an integral), which intuitively characterises the cost of passing the state.

Note that our budget analysis is an open framework within which an analyst could study various behaviour of insiders. To make this framework concrete, we give some examples of interesting analysis questions:

1) What is the maximum expected "cost" of the insider to launch an attack, e.g., eg `R"cost"min=?[F (loc=succ and tick=1)]`.? For this type of budget analysis, we shall use real-valued reachability reward query, which associates a cost with each path of the MDP. More specifically, they refer to the reward accumulated along a path until a certain point is reached.
2) Given the budget which the insider is willing to (or is able to) pay, what is the maximal success probability and the associated optimal strategy, *e.g.*, `Pmax=? [F <=x (loc=succ and tick=1) ]`.? For this type of question, it is assumed that the insider has a prescribed budget x, and the analyst is interested in computing the maximal success probability and the associated optimal strategy.
3) What is the strategy the insider might take under which a certain profit is guaranteed *and* the expected "cost" is affordable?
   To cope with such a question, we need a multi-objective query. First of all, apart from the rewards structure "cost" defined before, we introduce another rewards structure *profit* which specifies the gain of the insider. For instance,

```
rewards "profit"
    loc=6 & tick=1 : 2000;
    loc=4 & tick=1 : 1500;
    ...
endrewards
```

   We then can query PRISM to compute approximately the Pareto curve for this pair of objectives.

```
multi(R{"cost"}min=?[F loc=succ and tick=1],
R{"profit"} max= [F loc=succ and tick=1])
```

We briefly discuss why these questions trigger a more thorough analysis. For the first question, common sense shows a rational insider will not choose an attack where the expected cost exceeds some threshold. Hence, the organisation might not worry too much about such strategies as insiders are unlikely to follow them. For the second question, the analyst can estimate the risk and find the potential strategy of an insider for a given budget. For the third question, we postulate that the insider wants to maximise the profit by launching an insider attack while keeping the expected cost below a certain level. Such strategies should be the focus of the analyst and should be prevented by the organisation.

III. Conclusion

We have provided a quantitative analysis that enables security analysts to estimate the concrete insider risk given an infrastructure, the actors and their profiles. The framework identifies insiders' intention to attack using Bayesian networks followed by computing the probability of success for this actor with probabilistic model checking. It has been supported by Matlab and PRISM in the two steps respectively.

b) Related Work: We base our work on existing taxonomies of insiders [10], [2]. [4] uses Higher Order Logic to model insider threats enabling a possibilistic analysis without quantitatively estimating the risk of an insider attack. The Insider threat patterns provided by CERT [2] use System Dynamics models, but do not support probabilities quantifying these dependencies nor any of the probabilistic analysis that we propose here. Axelrad et al. [1] have used Bayesian networks for modelling insider threats, but did not provide analysis or implementation. [8] presents a methodology for insider threats assessment and mitigation. Bearing similar aims, the approaches are quite different: [8] is largely based on the ADVISE method [6]. The models therein lack probabilistic/cost aspects, and the analysis is based on simulation while our method is based on model checking. Very recently, [9] outlines a research proposal for studying security with "human in the loop".

For the future work, we are currently building up a tool chain which integrates different analysis presented in the paper into a monolithic tool. Moreover, the current framework considers MDPs. An interesting view is to treat the company and (potentially multiple) insiders as players of a *game*, which would allow analysing, for example, coalition (as the ambitious leader example shows) of insiders.

References

[1] E. T. Axelrad, P. J. Sticha, O. Brdiczka, J. Shen. A bayesian network model for predicting insider threats. In *IEEE Security and Privacy Workshops*,p. 82–89. IEEE, 2013.

[2] D. M. Cappelli, A. P. Moore, R. F. Trzeciak. *The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes*. Addison-Wesley, 2012.

[3] T. Chen, T. Han, F. Kammueller, I. Nemli, C. W. Probst. Model Based Analysis of Insider Threats. Available from http://www.eis.mdx.ac.uk/staffpages/taoluechen/

[4] F. Kammüller, C. W. Probst. Modeling and verification of insider threats using logical analysis. *IEEE Systems Journal*. To appear.

[5] M. Z. Kwiatkowska, G. Norman, D. Parker. PRISM 4.0: Verification of Probabilistic Real-time Systems. In *CAV11*, LNCS 6806, p585-591, 2011.

[6] E. LeMay, M. D. Ford, K. Keefe, W. H. Sanders, C. Muehrcke. Model-based security metrics using adversary view security evaluation (ADVISE). In *QEST'11*, p. 191–200. IEEE, 2011.

[7] K. Murphy. The bayes net toolbox for matlab. *Computing Science and Statistics*, 33(2):1024 – 1034, 2001.

[8] N. Nostro, A. Ceccarelli, A. Bondavalli, F. Brancati. Insider threat assessment: a model-based methodology. *Operating Systems Review*, 48(2):3–12, 2014.

[9] M. Noureddine, K. Keefe, W. H. Sanders, M. Bashir. Quantitative security metrics with human in the loop. In *HotSoS'15*, p. 21:1–21:2, 2015.

[10] J. R. C. Nurse, O. Buckley, P. A. Legg, M. Goldsmith, S. Creese, G. R. T. Wright, M. Whitty. Understanding Insider Threat: A Framework for Characterising Attacks. In *WRIT'14*. IEEE, 2014.